



Entre neutralidade e confronto: preconceitos de gênero em narrativas produzidas por inteligências artificiais generativas

Plínio Alexandre dos Santos Caetano, Larissa Caroline de Oliveira

Resumo: O avanço das inteligências artificiais generativas tem ampliado o uso de sistemas automatizados de produção textual em contextos educacionais, institucionais e comunicacionais, suscitando debates éticos acerca da neutralidade algorítmica e da reprodução de vieses sociais. Este artigo analisa como preconceitos de gênero se manifestam em narrativas produzidas pelo ChatGPT, à luz da Análise Crítica do Discurso e da Teoria da Performatividade de Gênero. Trata-se de uma pesquisa qualitativa e exploratória, baseada na análise de um corpus de respostas geradas a partir de *prompts* estruturados, elaborados para provocar reflexões sobre papéis sociais, competências e expectativas atribuídas a homens e mulheres. Os resultados indicam que, embora o sistema adote, em determinados contextos, discursos alinhados a princípios normativos de igualdade, persistem associações implícitas que reiteram estereótipos de gênero, especialmente nos eixos da liderança, das competências profissionais e da expressão emocional. Conclui-se que a produção textual automatizada não é neutra, mas atravessada por discursos socialmente situados, evidenciando a necessidade de abordagens críticas no uso educacional e no desenvolvimento de inteligências artificiais generativas mais éticas e inclusivas.

Palavras-chave: Inteligência artificial generativa; Performatividade de gênero; Discurso; Produção textual automatizada; Ética em tecnologia.

Abstract: The advancement of generative artificial intelligence has significantly expanded the use of automated text production systems in educational, institutional, and communicational contexts, raising ethical debates regarding algorithmic neutrality and the reproduction of social biases. This

article analyzes how gender prejudice manifests in narratives generated by ChatGPT, drawing on Critical Discourse Analysis and Gender Performativity Theory. The study adopts a qualitative and exploratory approach, based on the analysis of a corpus of responses generated from structured prompts designed to elicit reflections on social roles, competencies, and expectations attributed to men and women. The results indicate that, although the system adopts discourse aligned with normative principles of gender equality in certain contexts, implicit associations persist that reiterate gender stereotypes, particularly in relation to leadership, professional competencies, and emotional expression. It is concluded that automated text production is not neutral but embedded in socially situated discourses, highlighting the need for critical approaches to the educational use and development of more ethical and inclusive generative artificial intelligence systems.

Keywords: Generative artificial intelligence; Gender performativity; Discourse; Automated text production; Ethics in technology.

1. Introdução

O avanço das inteligências artificiais generativas tem transformado de maneira significativa os modos de produção, circulação e consumo de textos na sociedade contemporânea. Ferramentas baseadas em modelos de linguagem de grande escala passaram a ocupar espaços relevantes em atividades educacionais, profissionais e comunicacionais, reconfigurando práticas discursivas e relações sociais mediadas pela tecnologia.

Entre essas ferramentas, o ChatGPT destaca-se pela capacidade de gerar textos coesos e contextualmente adequados, simulando interações discursivas complexas e aproximando-se de padrões comunicativos humanos. Tal característica contribui para sua ampla adoção, mas também intensifica debates éticos sobre autoria, confiabilidade, neutralidade algorítmica e responsabilidade social no uso dessas tecnologias.

Embora frequentemente apresentadas como neutras e objetivas, as inteligências artificiais são treinadas a partir de grandes volumes de dados produzidos social e historicamente. Como resultado, esses sistemas tendem a assimilar valores, hierarquias e preconceitos presentes nos contextos socioculturais que alimentam seus bancos de dados, reproduzindo-os, ainda que de forma não intencional.

Estudos recentes têm evidenciado que modelos de linguagem podem reiterar estereótipos relacionados a gênero, raça e classe social, revelando limites importantes no discurso da

neutralidade tecnológica (RIBEIRO; CORDEIRO; FUMACH, 2022; THE LANCET DIGITAL HEALTH, 2023). Essas reproduções simbólicas tornam-se especialmente problemáticas quando tais sistemas passam a influenciar processos educativos e formativos.

No que se refere às relações de gênero, a linguagem ocupa papel central na manutenção ou na contestação de estruturas sociais desiguais. Discursos que associam determinadas competências, comportamentos ou papéis sociais a homens ou mulheres contribuem para a naturalização de hierarquias historicamente construídas.

A Análise Crítica do Discurso, conforme proposta por Fairclough (2020), permite compreender como práticas discursivas não apenas refletem a realidade social, mas também a produzem e a legitimam. Nesse sentido, textos gerados por inteligências artificiais podem ser analisados como práticas sociais que participam da reprodução ou da transformação de ideologias dominantes.

De forma complementar, a Teoria da Performatividade de Gênero, desenvolvida por Judith Butler, compreende o gênero como uma construção discursiva e performativa, reiterada por meio de práticas simbólicas cotidianas. Essa perspectiva oferece subsídios teóricos para analisar como narrativas automatizadas podem reforçar ou tensionar normas de gênero socialmente instituídas.

A articulação entre inteligência artificial, discurso e gênero torna-se, portanto, um campo relevante de investigação, especialmente diante da crescente incorporação dessas tecnologias em contextos educacionais e institucionais. Analisar criticamente tais produções é fundamental para compreender seus impactos éticos, pedagógicos e sociais.

Nesse contexto, este artigo teve como objetivo geral analisar como preconceitos de gênero se manifestam em narrativas produzidas pelo ChatGPT, a partir de uma abordagem fundamentada na Análise Crítica do Discurso e na Teoria da Performatividade de Gênero.

Como objetivos específicos, buscou-se: identificar padrões discursivos associados a estereótipos de gênero nas respostas geradas pela inteligência artificial; comparar produções textuais obtidas sob diferentes configurações de interação; e refletir sobre as implicações éticas e educacionais dessas narrativas no uso de tecnologias baseadas em inteligência artificial.

Do ponto de vista metodológico, a pesquisa caracteriza-se como qualitativa e exploratória, voltada à compreensão aprofundada dos sentidos produzidos nos textos analisados. Essa abordagem permite examinar nuances discursivas e ideológicas que não seriam captadas por métodos quantitativos.

O *corpus* da pesquisa é composto por respostas geradas pelo ChatGPT a partir de prompts previamente estruturados, elaborados para provocar reflexões relacionadas a gênero, papéis sociais

e competências atribuídas a homens e mulheres. As respostas foram organizadas e sistematizadas para fins de análise comparativa.

A análise dos dados foi realizada à luz dos pressupostos da Análise Crítica do Discurso, considerando aspectos como conteúdo, tom e implicações sociais, articulados à noção de performatividade de gênero proposta por Butler. Essa combinação teórica possibilita compreender como discursos automatizados podem operar na reprodução de normas sociais.

Ao adotar esse enquadramento analítico, o estudo busca ultrapassar uma leitura meramente técnica da inteligência artificial, situando-a como prática social inserida em relações de poder e disputas simbólicas. Tal perspectiva contribui para o debate crítico sobre tecnologia, linguagem e desigualdade.

Espera-se que os resultados desta investigação contribuam para a reflexão sobre o desenvolvimento e o uso de inteligências artificiais mais éticas, inclusivas e socialmente responsáveis, especialmente no campo educacional, onde a mediação tecnológica exerce papel crescente na formação de sujeitos e discursos.

2. Inteligência artificial, linguagem e vieses sociais

A inteligência artificial (IA) tem sido progressivamente incorporada a diferentes esferas da vida social, especialmente por meio de sistemas capazes de processar linguagem natural e produzir textos de forma automatizada. Esses sistemas não apenas executam tarefas técnicas, mas participam ativamente da mediação simbólica entre sujeitos, instituições e discursos.

Modelos de linguagem de grande escala, como o ChatGPT, operam a partir de arquiteturas estatísticas complexas treinadas com vastos conjuntos de dados textuais disponíveis na internet. Esses dados refletem práticas discursivas socialmente situadas, atravessadas por valores culturais, ideologias e relações de poder historicamente construídas (GOMES, 2010; KRETTEK, 2023).

Embora frequentemente apresentados como ferramentas neutras, os sistemas de inteligência artificial não são isentos de vieses. Ao contrário, estudos têm demonstrado que algoritmos de aprendizagem de máquina tendem a reproduzir padrões discriminatórios presentes nos dados de treinamento, especialmente quando esses dados são marcados por desigualdades sociais estruturais (RIBEIRO; CORDEIRO; FUMACH, 2022).

No campo da linguagem, tais vieses manifestam-se por meio da escolha lexical, da organização sintática e da construção de sentidos que reforçam estereótipos ou hierarquias sociais. A produção

textual automatizada, nesse sentido, não se limita à geração de frases gramaticalmente corretas, mas envolve processos de significação que dialogam com discursos socialmente legitimados.

A Análise Crítica do Discurso contribui para a compreensão desse fenômeno ao conceber a linguagem como prática social, indissociável das condições históricas e ideológicas de sua produção. Para Fairclough (2020), os discursos não apenas refletem a realidade social, mas atuam na sua constituição, legitimando ou contestando estruturas de poder.

Ao aplicar essa perspectiva à inteligência artificial, torna-se possível compreender os textos gerados por sistemas automatizados como produtos discursivos que participam da circulação de ideologias. Assim, respostas produzidas por modelos de linguagem podem reforçar visões de mundo hegemônicas ou, em determinados contextos, tensioná-las.

A noção de viés algorítmico emerge, portanto, como elemento central nas discussões contemporâneas sobre ética em inteligência artificial. Esse conceito refere-se à tendência de sistemas automatizados apresentarem resultados sistematicamente desfavoráveis a determinados grupos sociais, reproduzindo desigualdades de gênero, raça ou classe (THE LANCET DIGITAL HEALTH, 2023).

No caso específico das questões de gênero, a linguagem desempenha papel fundamental na construção e manutenção de expectativas sociais sobre comportamentos, competências e papéis atribuídos a homens e mulheres. Quando tais expectativas são incorporadas por sistemas automatizados, corre-se o risco de amplificar discursos discriminatórios sob a aparência de neutralidade tecnológica.

Pesquisas recentes têm alertado que, mesmo quando modelos de linguagem adotam um tom aparentemente inclusivo, podem reproduzir vieses de forma implícita, por meio de associações sutis entre gênero e determinadas características sociais ou profissionais (TRIGO, 2024).

A naturalização desses discursos é particularmente preocupante em contextos educacionais, nos quais a inteligência artificial passa a ser utilizada como ferramenta de apoio ao ensino, à escrita e à produção de conhecimento. Nesses espaços, a reprodução de estereótipos pode impactar processos formativos e contribuir para a manutenção de desigualdades simbólicas.

A discussão sobre ética em inteligência artificial, portanto, ultrapassa questões técnicas e envolve reflexões sobre responsabilidade social, transparência algorítmica e justiça cognitiva. Desenvolver sistemas mais equitativos implica reconhecer que tecnologia e sociedade estão profundamente interligadas.

Nesse sentido, torna-se fundamental analisar criticamente não apenas os resultados produzidos pelas inteligências artificiais, mas também os discursos que legitimam seu uso acrítico. A ideia de eficiência tecnológica não pode se sobrepor à necessidade de avaliação ética de seus impactos sociais.

A literatura especializada aponta para a importância de diversificar os dados de treinamento, ampliar a participação humana nos processos de validação e incorporar perspectivas interdisciplinares no desenvolvimento de sistemas de IA (RIBEIRO; CORDEIRO; FUMACH, 2022).

Ao articular inteligência artificial, linguagem e vieses sociais, este estudo posiciona-se no campo das ciências humanas aplicadas à tecnologia, reconhecendo a IA como prática discursiva inserida em relações de poder e disputas simbólicas.

Essa abordagem teórica oferece subsídios para a análise das narrativas produzidas pelo ChatGPT, permitindo compreender de que maneira a produção textual automatizada pode atuar tanto na reprodução quanto na problematização de desigualdades sociais, especialmente no que se refere às relações de gênero.

3. Gênero, discurso e performatividade na produção textual automatizada

A compreensão das relações de gênero no campo da linguagem exige o reconhecimento de que discursos não são neutros, mas atravessados por normas sociais, históricas e culturais que regulam identidades e papéis sociais. Nesse sentido, a produção textual, seja humana ou automatizada, participa ativamente da construção e da manutenção de sentidos sobre o que se entende por masculino, feminino e outras expressões de gênero.

A Teoria da Performatividade de Gênero, desenvolvida por Judith Butler, concebe o gênero não como uma essência biológica, mas como uma construção discursiva reiterada por meio de práticas simbólicas e sociais. Para a autora, o gênero é continuamente produzido e reproduzido por atos performativos que, ao se repetirem, naturalizam determinadas normas e expectativas sociais (BUTLER, 2019).

Essa perspectiva teórica permite compreender que narrativas, ao associarem comportamentos, competências ou emoções a determinados gêneros, contribuem para a consolidação de padrões normativos. Tais padrões tornam-se especialmente relevantes quando incorporados por tecnologias que produzem textos em larga escala, como os sistemas de inteligência artificial generativa.

No contexto da produção textual automatizada, modelos de linguagem operam a partir de regularidades estatísticas extraídas de grandes bases de dados. Como esses dados são compostos por discursos socialmente situados, as normas de gênero presentes na sociedade tendem a ser incorporadas, ainda que de forma implícita, às respostas geradas por esses sistemas.

A Análise Crítica do Discurso contribui para a compreensão desse fenômeno ao evidenciar como a linguagem atua na legitimação de relações de poder. Fairclough (2020) destaca que discursos não apenas refletem estruturas sociais, mas também participam de sua reprodução. Assim, textos gerados por inteligências artificiais podem reforçar hierarquias de gênero ao reiterarem associações naturalizadas entre gênero e determinados papéis sociais.

A articulação entre a Análise Crítica do Discurso e a teoria de Butler possibilita compreender como a performatividade de gênero se manifesta em produções textuais automatizadas. Quando um sistema associa liderança, racionalidade ou autoridade a atributos masculinos, ou sensibilidade e cuidado a atributos femininos, está reproduzindo performances de gênero socialmente reguladas.

Mesmo quando essas associações são apresentadas de forma crítica ou aparentemente neutra, sua enunciação pode contribuir para a manutenção de normas discursivas hegemônicas. Butler argumenta que a simples repetição de categorias normativas, ainda que com intenção de desconstrução, pode reforçar sua presença no imaginário social (BUTLER, 2019).

Nesse sentido, a linguagem gerada por inteligências artificiais não deve ser analisada apenas pelo conteúdo explícito, mas também pelos pressupostos e implícitos que estruturam suas narrativas. A escolha lexical, o tom discursivo e as metáforas utilizadas são elementos que revelam modos específicos de performar o gênero no discurso.

A produção textual automatizada assume relevância particular em contextos educacionais e institucionais, nos quais essas tecnologias passam a ser utilizadas como mediadoras do conhecimento. Nesses espaços, a reprodução de estereótipos de gênero pode impactar processos formativos e contribuir para a naturalização de desigualdades simbólicas.

Pesquisas no campo dos estudos de gênero têm alertado que a tecnologia, longe de ser neutra, pode reforçar estruturas patriarcais quando desenvolvida e utilizada sem uma reflexão crítica sobre seus fundamentos sociais (TRIGO, 2024). A inteligência artificial, nesse cenário, torna-se um novo espaço de disputa simbólica.

A noção de performatividade aplicada à inteligência artificial permite compreender que sistemas automatizados não apenas reproduzem discursos existentes, mas também participam da circulação e

da legitimação de determinadas visões de mundo. Ao gerar textos que reiteram normas de gênero, esses sistemas contribuem para a manutenção de práticas discursivas excludentes.

Por outro lado, a análise crítica dessas produções abre possibilidades para a problematização e a transformação desses discursos. Ao tornar visíveis os mecanismos pelos quais a performatividade de gênero se manifesta na produção textual automatizada, cria-se espaço para o desenvolvimento de tecnologias mais conscientes e inclusivas.

A literatura especializada aponta para a necessidade de incorporar perspectivas feministas e críticas no desenvolvimento e na avaliação de sistemas de inteligência artificial, de modo a mitigar a reprodução de estereótipos e ampliar a diversidade de representações (RIBEIRO; CORDEIRO; FUMACH, 2022).

Assim, a análise das narrativas produzidas pelo ChatGPT, à luz da performatividade de gênero, permite compreender como a inteligência artificial se insere em dinâmicas sociais mais amplas, evidenciando que a tecnologia não apenas reflete a sociedade, mas também participa ativamente da construção de sentidos sobre gênero.

Essa abordagem teórica fundamenta a análise empírica apresentada neste estudo, contribuindo para o debate sobre o uso ético da inteligência artificial e para a promoção de práticas discursivas mais equitativas no campo da produção textual automatizada.

4. Metodologia

A presente pesquisa caracteriza-se como um estudo de natureza qualitativa, de caráter exploratório e interpretativo, adequado à análise de fenômenos sociais complexos mediados pela linguagem. A abordagem qualitativa permite compreender sentidos, significados e construções discursivas, priorizando a profundidade analítica em detrimento da mensuração numérica (MINAYO, 2014; GIL, 2019).

Quanto aos objetivos, a pesquisa é classificada como exploratória, uma vez que busca aprofundar a compreensão de um fenômeno contemporâneo ainda em consolidação no campo científico, a saber, a manifestação de preconceitos de gênero em narrativas produzidas por inteligências artificiais generativas. Esse tipo de delineamento é indicado quando o tema demanda maior familiaridade teórica e empírica, possibilitando a formulação de interpretações críticas (GIL, 2019).

Do ponto de vista epistemológico, o estudo insere-se no campo das ciências humanas aplicadas à tecnologia, compreendendo a linguagem como prática social e a inteligência artificial como artefato

sociotécnico situado histórica e culturalmente. Tal perspectiva reconhece que tecnologias não são neutras, mas atravessadas por valores, ideologias e relações de poder (FEENBERG, 2010).

O objeto de análise consiste em narrativas textuais produzidas pelo ChatGPT, modelo de linguagem baseado em arquiteturas de aprendizagem profunda e amplamente utilizado em contextos educacionais e comunicacionais. A escolha desse sistema justifica-se por sua relevância social e pelo impacto crescente de suas produções discursivas na mediação do conhecimento.

A constituição do corpus ocorreu a partir da elaboração de prompts estruturados, formulados com o objetivo de provocar respostas relacionadas a papéis sociais, competências, emoções e expectativas associadas a homens e mulheres. A definição criteriosa dos comandos segue orientações metodológicas para pesquisas qualitativas baseadas em documentos e textos, nas quais a seleção do material empírico deve estar alinhada aos objetivos analíticos do estudo (CELLARD, 2008).

As interações com o sistema foram realizadas sob diferentes configurações de resposta, permitindo a comparação entre narrativas geradas em contextos de maior controle normativo e aquelas produzidas em condições discursivas menos restritivas. Esse procedimento possibilitou identificar variações no tom discursivo, na escolha lexical e na explicitação de estereótipos de gênero.

As respostas obtidas foram organizadas em um banco de dados textual, preservando-se a integridade das produções originais. Posteriormente, realizou-se a seleção de excertos considerados mais representativos, adotando-se critérios de recorrência temática, relevância discursiva e potencial analítico, conforme recomenda Bardin (2016) para estudos qualitativos baseados em material textual.

A análise dos dados foi conduzida à luz da Análise Crítica do Discurso, conforme proposta por Fairclough (2020), que compreende o discurso como prática social articulada a relações de poder e ideologia. Foram observados elementos como conteúdo temático, estratégias linguísticas, tom discursivo e implicações sociais das narrativas produzidas.

De forma complementar, os dados foram interpretados a partir da Teoria da Performatividade de Gênero, desenvolvida por Judith Butler, que concebe o gênero como uma construção discursiva reiterada por práticas simbólicas e sociais. Essa abordagem permitiu identificar como normas de gênero são performadas, naturalizadas ou tensionadas nas produções textuais automatizadas (BUTLER, 2019).

A articulação entre a Análise Crítica do Discurso e a teoria da performatividade possibilitou uma leitura integrada dos dados, considerando tanto os mecanismos discursivos de reprodução

ideológica quanto os processos simbólicos que sustentam expectativas sociais de gênero. Tal combinação metodológica é coerente com estudos interdisciplinares que analisam linguagem, poder e tecnologia.

O procedimento analítico privilegiou a interpretação qualitativa e contextualizada dos sentidos produzidos, evitando generalizações estatísticas. O foco esteve na compreensão dos padrões discursivos recorrentes e de suas possíveis implicações sociais e educacionais, conforme defendem Minayo (2014) e Fairclough (2020).

Do ponto de vista ético, a pesquisa não envolveu participantes humanos diretos, uma vez que o corpus analisado é composto exclusivamente por textos gerados por um sistema de inteligência artificial. Ainda assim, adotou-se uma postura ética reflexiva, considerando os impactos sociais do uso e da análise dessas tecnologias, especialmente em contextos educacionais (THE LANCET DIGITAL HEALTH, 2023).

As limitações do estudo relacionam-se ao recorte do corpus e às especificidades do sistema analisado, não se pretendendo extrapolar os resultados para todos os modelos de linguagem existentes. Contudo, os achados oferecem subsídios relevantes para reflexões mais amplas sobre inteligência artificial, linguagem e gênero.

Por fim, o delineamento metodológico adotado busca contribuir para o fortalecimento de abordagens críticas no estudo da inteligência artificial, evidenciando a importância de análises interdisciplinares que articulem tecnologia, discurso e questões sociais, fundamentando a análise dos resultados apresentada na seção seguinte.

5. Resultados e discussão

A análise das narrativas produzidas pelo ChatGPT evidenciou que a produção textual automatizada não é homogênea, variando conforme as configurações de interação e os estímulos discursivos apresentados. Ainda assim, foi possível identificar padrões recorrentes que revelam a presença de estereótipos e normas de gênero, ora explicitados, ora atenuados por estratégias discursivas de neutralização. Os resultados são apresentados a seguir, organizados em eixos analíticos.

5.1 Gênero e liderança

No eixo referente à liderança, observou-se que as respostas geradas pelo sistema tendem a adotar, em configurações mais normativas, um discurso alinhado à igualdade formal entre gêneros, enfatizando competências individuais e rejeitando associações diretas entre liderança e gênero. Esse

posicionamento dialoga com discursos institucionais contemporâneos que promovem a equidade como valor organizacional (FAIRCLOUGH, 2020).

Entretanto, em configurações menos restritivas, emergiram associações implícitas que vinculam liderança masculina a atributos como autoridade, racionalidade e firmeza, enquanto a liderança feminina foi frequentemente associada à sensibilidade, empatia e inspiração. Tais associações refletem performances de gênero socialmente naturalizadas, nas quais o masculino ocupa posições de comando e o feminino é deslocado para papéis de mediação emocional (BUTLER, 2019).

Esse resultado indica que, mesmo quando o discurso se apresenta como inclusivo, a repetição de atributos generificados contribui para a manutenção de hierarquias simbólicas. Conforme argumenta Butler (2019), a performatividade de gênero opera justamente pela reiterada associação entre determinados traços e identidades, o que pode ser reproduzido por sistemas automatizados.

5.2 Competências profissionais e papéis sociais

No eixo das competências profissionais, as narrativas analisadas revelaram uma tendência à reprodução de estereótipos associados à divisão sexual do trabalho. Em algumas respostas, habilidades técnicas e lógicas foram associadas com maior frequência aos homens, enquanto habilidades comunicacionais e de cuidado foram atribuídas às mulheres, ainda que acompanhadas de ressalvas críticas.

Esse movimento discursivo ambíguo evidencia o que Fairclough (2020) denomina como estratégias de mitigação ideológica, nas quais o discurso reconhece a existência de desigualdades, mas simultaneamente as reproduz ao reiterar categorias normativas. A menção crítica aos estereótipos não impede que estes sejam reativados no próprio ato enunciativo.

Do ponto de vista da performatividade, a enumeração de competências generificadas reforça expectativas sociais que regulam trajetórias profissionais, contribuindo para a naturalização de desigualdades no mercado de trabalho. A inteligência artificial, ao reproduzir essas associações, atua como agente de circulação desses discursos, ampliando seu alcance simbólico.

5.3 Emoções, racionalidade e legitimidade discursiva

Outro eixo analítico relevante refere-se à relação entre gênero e expressão emocional. As narrativas produzidas pelo sistema indicaram que a expressão de emoções por mulheres em posições de liderança tende a ser apresentada como elemento problemático ou passível de julgamento, enquanto comportamentos semelhantes em homens são frequentemente naturalizados ou relativizados.

Esse padrão discursivo dialoga com estudos que apontam a penalização simbólica das mulheres quando estas transgridem normas de contenção emocional associadas à liderança (TRIGO, 2024). A linguagem gerada pelo sistema reproduz, assim, expectativas normativas que regulam a legitimidade discursiva de homens e mulheres em espaços de poder.

Sob a ótica da Análise Crítica do Discurso, tais narrativas reforçam estruturas sociais que associam racionalidade ao masculino e emocionalidade ao feminino, legitimando desigualdades de gênero por meio de práticas linguísticas aparentemente neutras (FAIRCLOUGH, 2020).

5.4 Igualdade de gênero e neutralidade discursiva

No eixo relacionado à igualdade de gênero, observou-se que o ChatGPT frequentemente adota um discurso alinhado a princípios normativos de equidade, destacando benefícios sociais e organizacionais da igualdade. Esse posicionamento reflete a incorporação de discursos progressistas amplamente difundidos em políticas institucionais e documentos oficiais.

Todavia, a análise crítica revela que essa neutralidade discursiva pode operar como forma de silenciamento das desigualdades estruturais. Ao tratar a igualdade como um dado ou como um objetivo consensual, o discurso tende a minimizar conflitos e a invisibilizar assimetrias persistentes, conforme alertam estudos críticos sobre tecnologia e justiça social (THE LANCET DIGITAL HEALTH, 2023).

Nesse sentido, a neutralidade algorítmica apresenta-se como estratégia discursiva que, embora evite enunciados explicitamente discriminatórios, não enfrenta de forma substantiva as bases sociais da desigualdade de gênero.

5.5 Autorreflexividade e reconhecimento de vieses

Por fim, no eixo da autorreflexividade, as respostas do sistema demonstraram uma postura ambígua diante da possibilidade de reprodução de preconceitos. Em algumas narrativas, há reconhecimento explícito da existência de vieses nos dados de treinamento e do risco de respostas tendenciosas; em outras, observa-se uma negação ou relativização desses vieses.

Esse movimento discursivo evidencia tensões entre transparência e legitimação tecnológica. Conforme apontam Ribeiro, Cordeiro e Fumach (2022), o reconhecimento parcial dos limites dos sistemas de inteligência artificial pode funcionar tanto como mecanismo de responsabilidade ética quanto como estratégia de preservação da credibilidade tecnológica.

A análise desses resultados reforça a compreensão de que a inteligência artificial não apenas reflete discursos sociais, mas participa ativamente de sua circulação e legitimação. Ao reproduzir normas

performativas de gênero, ainda que de forma mitigada, sistemas como o ChatGPT contribuem para a manutenção de desigualdades simbólicas, evidenciando a necessidade de abordagens críticas e interdisciplinares no desenvolvimento e uso dessas tecnologias.

De modo geral, a análise dos eixos evidenciou que as narrativas produzidas pelo ChatGPT oscilam entre discursos alinhados a princípios normativos de igualdade de gênero e a reprodução, ainda que mitigada, de estereótipos socialmente naturalizados. Essa ambivalência discursiva revela que a produção textual automatizada não rompe plenamente com estruturas simbólicas consolidadas, mas tende a reatualizá-las sob a aparência de neutralidade tecnológica. Conforme discutido ao longo desta seção, tais padrões se manifestam de forma diferenciada nos eixos relacionados à liderança, às competências profissionais, à expressão emocional, à igualdade de gênero e à autorreflexividade do sistema diante de seus próprios vieses. Com o objetivo de sistematizar os principais achados e explicitar a articulação entre padrões discursivos, referenciais teóricos e implicações sociais, apresenta-se o Quadro 1, que sintetiza os eixos analíticos e os elementos centrais identificados na análise das narrativas.

Quadro 1 – Síntese dos eixos analíticos, padrões discursivos e implicações sociais nas narrativas geradas pelo ChatGPT

Eixo analítico	Padrões discursivos identificados	Referenciais teóricos mobilizados	Implicações críticas
Gênero e liderança	Discurso formalmente igualitário em contextos normativos; associação implícita entre liderança masculina e autoridade, e liderança feminina e empatia em contextos menos restritivos.	Fairclough (2020); Butler (2019)	Reforço simbólico de hierarquias de gênero, mesmo sob aparência de neutralidade; manutenção de performances tradicionais de liderança.
Competências profissionais e papéis	Associação recorrente de competências técnicas e racionais ao masculino e	Fairclough (2020); Butler (2019); Trigo	Naturalização da divisão sexual do trabalho e reprodução de expectativas

Eixo analítico	Padrões discursivos identificados	Referenciais teóricos mobilizados	Implicações críticas
sociais	competências comunicacionais e de cuidado ao feminino, ainda que acompanhadas de ressalvas críticas.	(2024)	sociais que regulam trajetórias profissionais.
Emoções, racionalidade e legitimidade discursiva	Penalização simbólica da expressão emocional feminina em posições de liderança; tolerância maior à expressão emocional masculina.	Butler (2019); Trigo (2024)	Reforço da dicotomia razão/emoção associada a gênero, legitimando desigualdades no acesso a posições de poder.
Igualdade de gênero e neutralidade discursiva	Adoção de discurso institucional progressista, com foco em equidade e benefícios organizacionais, minimizando conflitos estruturais.	Fairclough (2020); The Lancet Digital Health (2023)	Neutralidade discursiva como estratégia de silenciamento das desigualdades estruturais de gênero.
Autorreflexividade e reconhecimento de vieses	Reconhecimento parcial dos vieses algorítmicos; alternância entre transparência ética e relativização dos limites do sistema.	Ribeiro, Cordeiro e Fumach (2022); Fairclough (2020)	Tensão entre responsabilidade ética e legitimação tecnológica; manutenção da credibilidade do sistema.

Fonte: Elaboração própria, a partir da análise das narrativas produzidas pelo ChatGPT (2024).

6. Considerações finais

O presente estudo teve como objetivo geral analisar como preconceitos de gênero se manifestam em narrativas produzidas pelo ChatGPT, à luz da Análise Crítica do Discurso e da Teoria da Performatividade de Gênero. A partir da análise qualitativa das respostas geradas pelo sistema, foi possível evidenciar que a produção textual automatizada não se configura como neutra, mas

atravessada por discursos socialmente situados, que refletem e, em determinados contextos, reforçam normas e hierarquias de gênero.

No que se refere aos objetivos específicos, o primeiro — identificar padrões discursivos associados a estereótipos de gênero — foi alcançado ao se observar recorrências simbólicas nas narrativas relacionadas à liderança, às competências profissionais e à expressão emocional. Mesmo quando acompanhadas de ressalvas críticas ou de um discurso formalmente igualitário, essas associações reiteram expectativas sociais historicamente construídas.

O segundo objetivo específico, que consistiu em comparar produções textuais geradas sob diferentes configurações de interação, permitiu evidenciar variações significativas no tom discursivo e na explicitação dos estereótipos. Enquanto respostas mais normativas tendem a adotar uma linguagem alinhada a discursos institucionais de equidade, configurações menos restritivas revelam vieses de gênero de forma mais explícita, indicando limites importantes da neutralidade algorítmica.

O terceiro objetivo específico — refletir sobre as implicações éticas e educacionais dessas narrativas — foi contemplado por meio da análise crítica dos impactos simbólicos da produção textual automatizada. Observou-se que a incorporação dessas tecnologias em contextos educacionais exige cautela, uma vez que discursos aparentemente neutros podem contribuir para a naturalização de desigualdades de gênero.

A síntese dos eixos analíticos demonstrou que a inteligência artificial generativa opera em um campo discursivo ambivalente, no qual coexistem tentativas de alinhamento a valores progressistas e a reiteração de performances de gênero normativas. Essa ambivalência reforça a compreensão de que tecnologias não apenas refletem a sociedade, mas participam ativamente da circulação e legitimação de discursos.

Do ponto de vista teórico, a articulação entre a Análise Crítica do Discurso e a Teoria da Performatividade de Gênero mostrou-se produtiva para compreender a produção textual automatizada como prática social. Essa combinação permitiu evidenciar que a performatividade de gênero não se limita às interações humanas, mas também se manifesta em sistemas automatizados, ampliando o escopo dos estudos críticos sobre linguagem e tecnologia.

Como contribuição prática, a pesquisa oferece subsídios para a reflexão sobre o desenvolvimento e o uso responsável de inteligências artificiais generativas. A identificação de padrões discursivos generificados reforça a necessidade de diversificação dos dados de treinamento, de maior

transparência algorítmica e da incorporação de equipes interdisciplinares e diversas nos processos de desenvolvimento dessas tecnologias.

No campo educacional, os resultados indicam que o uso de ferramentas como o ChatGPT deve ser acompanhado de práticas pedagógicas críticas, que estimulem a leitura reflexiva das produções automatizadas. A mediação docente torna-se fundamental para problematizar vieses, promover a educação midiática e fortalecer a formação ética e cidadã dos estudantes.

É importante reconhecer, entretanto, os limites deste estudo. Por tratar-se de uma pesquisa de iniciação científica, o recorte empírico é restrito e circunscrito a um único sistema de inteligência artificial, não se pretendendo generalizar os resultados para todos os modelos de linguagem existentes. Além disso, o corpus analisado reflete um conjunto específico de prompts e configurações de interação.

Outro limite refere-se ao caráter qualitativo e interpretativo da análise, que privilegia a profundidade dos sentidos em detrimento da abrangência estatística. Ainda assim, essa escolha metodológica é coerente com os objetivos do estudo e com a natureza do fenômeno investigado.

Como agenda futura de pesquisa, destaca-se a necessidade de ampliar o escopo analítico para outros modelos de inteligência artificial generativa, bem como de realizar estudos comparativos entre diferentes plataformas e versões de sistemas de linguagem. Investigações longitudinais também podem contribuir para compreender como os discursos produzidos por essas tecnologias se transformam ao longo do tempo.

Sugere-se, ainda, o aprofundamento das análises interseccionais, incorporando categorias como raça, classe social e territorialidade, de modo a compreender de forma mais abrangente os impactos sociais da produção textual automatizada. Tais abordagens são especialmente relevantes diante da crescente presença das inteligências artificiais nos processos educativos e institucionais.

Por fim, este estudo reafirma a importância de abordagens críticas e interdisciplinares no debate sobre inteligência artificial generativa. Ao evidenciar que a produção textual automatizada está imersa em relações de poder e disputas simbólicas, contribui-se para a construção de um campo de pesquisa comprometido com a promoção da equidade, da ética e da responsabilidade social no uso das tecnologias emergentes.

Referências

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2016.

- BUTLER, Judith. **Problemas de gênero: feminismo e subversão da identidade**. 17. ed. Rio de Janeiro: Civilização Brasileira, 2019.
- CELLARD, André. A análise documental. In: POUPART, Jean et al. **A pesquisa qualitativa: enfoques epistemológicos e metodológicos**. Petrópolis: Vozes, 2008. p. 295–316.
- FAIRCLOUGH, Norman. **Discurso e mudança social**. 2. ed. Brasília: Editora Universidade de Brasília, 2020.
- FEENBERG, Andrew. **Questioning technology**. London: Routledge, 2010.
- GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 7. ed. São Paulo: Atlas, 2019.
- GOMES, Denilson Silva. **Inteligência artificial: conceitos e aplicações**. Revista Olhar Científico, Ariquemes, v. 1, n. 2, p. 1–14, ago./dez. 2010. Disponível em: https://www.professores.uff.br/screspo/wp-content/uploads/sites/127/2017/09/ia_intro.pdf. Acesso em: 12 maio 2023.
- KRETTEK, Christian. **ChatGPT: milestone-text-KI mit game-changer-potenzial**. Die Unfallchirurgie, v. 126, n. 3, p. 252–254, 2023.
- MINAYO, Maria Cecília de Souza. **O desafio do conhecimento: pesquisa qualitativa em saúde**. 14. ed. São Paulo: Hucitec, 2014.
- RIBEIRO, Gustavo A. M.; CORDEIRO, Paula I. R. V.; FUMACH, Daniel M. **O malware como meio de obtenção de prova e sua implementação no ordenamento jurídico brasileiro**. Revista Brasileira de Direito Processual Penal, Porto Alegre, v. 8, n. 3, p. 1463–1500, set./dez. 2022. Disponível em: <https://www.scielo.br/j/rbdpp/a/rhHb6tynNX5rNH74mNGHSrj/>. Acesso em: 12 maio 2023.
- THE LANCET DIGITAL HEALTH. **ChatGPT: friend or foe?** The Lancet Digital Health, London, v. 5, n. 3, p. e102–e103, 2023.
- TRIGO, Luiz Gonzaga Godoi. **As milenares origens do preconceito de gênero**. São Paulo: Universidade de São Paulo, 2024.

Autores:

Plínio Alexandre dos Santos Caetano

Doutor em Desenvolvimento Territorial e Meio Ambiente. Professor do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Campus Sertãozinho. Atua nas áreas de

educação, gestão e tecnologias educacionais; desenvolvendo atividades de gestão acadêmica, ensino, pesquisa e extensão, orientando projetos de iniciação científica e trabalhos acadêmicos.

ORCID: <https://orcid.org/0000-0003-1405-7553>.

Larissa Caroline de Oliveira

Estudante de graduação vinculada ao curso de Letras – Português Inglês do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) – Campus Sertãozinho. Bolsista de Iniciação Científica nos anos de 2024 e 2025, com atuação em pesquisas voltadas à inteligência artificial, análise do discurso, gênero e educação. Possui interesse em estudos interdisciplinares que articulam tecnologia, linguagem e questões sociais.